Network Intrusion Detection Model based on Fuzzy Support Vector Machine

Yanjun Long
Network Information Center of Yongzhou Vocational Technical College, Yongzhou Hunan, China
Email: zgledu@163.com

Jianquan Ouyang
College of Information Engineering of Xiangtan University, Xiangtan Hunan, China
Email: oyjq@xtu.edu.cn

Xinwen Sun
Network Information Center of Yongzhou Vocational Technical College, Yongzhou Hunan, China
Email:sxw_980318@163.com

Abstract—Network intrusion detection is of great importance in the research field of information security in computer networks. In this paper, we concentrate on how to automatically detect the network intrusion behavior utilizing fuzzy support vector machine. After analyzing the related works of the proposed paper, we introduce the main characteristics of fuzzy support vector machine, and demonstrate its formal description in detail. Next, the proposed intrusion detection system is organized as five modules, which are Data source, AAA protocol, FSVM module located in local computer, Guest computer and Terminals. Particularly, the intrusion detection module is constructed by four sections, which are data gathering section, data pre-processing section, intrusion detecting section and decision response section. Then, the intrusion detection algorithm based on fuzzy support vector machine is implemented by training process and testing process. Utilizing this algorithm, a sample in testing data can be judged whether it is belonged to network intrusion behavior. Finally, experimental results verify the effectiveness of our method comparing with other methods under different metric.

Index Terms—Intrusion Detection, Fuzzy Support Vector Machine, kernel function, Fuzzy membership

I. INTRODUCTION

Network technologies along with the development of Internet have greatly promoted the development of social information in people's daily life, including work and learning activities. As Internet has brought great convenience to people, Internet computer network information security has become increasingly more and more important. The reason lies in that the computer network is designed to improve cyber source more open to more people, and with little considering the data safety. With the popularity of the network, the problem of network security becomes more and more important. According to the authoritative organization statistics, there are billions of dollars lost in each year due to hacker attacking. Particularly, some key business on the Internet more and more possible to be attacked. As is showing in the Spafford report, there are 99% large companies which had happened large network intrusion events. Computer network security is an international problem, and Internet security system is damaged and caused economic losses up to hundreds of billions dollars in each year [1].

Intrusion detection is a technology of information security for intrusion detection in computer networks, which is the cornerstone of network information security active defense technology. Aiming at the distributed, more and more frequent multi-objective multi-stage network attack, and the unknown security issues which could be happened in the next generation Internet, it is of great importance to improve the detection efficiency of intrusion detection system and. The machine learning method is widely used for classification and prediction, particularly; machine learning has recently been used in intrusion detection field. However, these methods such as correlation between samples, some problems such as repeat the training sample, long training time and intrusion sample labeling problem have not been solved successfully.

In 1980, James Anderson firstly put forward the basic concept of intrusion detection in the technical report titled “Computer Security Threat Monitoring and Surveillance”, of which the authors present a unified architecture for the intrusion and intrusion detection firstly. In this report, the concept of “Intrusion” is defined as the unauthorized access or operation under the condition of information, which could cause an attempt to the system unreliable or unavailable. Meanwhile, the threats related to computer systems are divided into three categories, including: external infiltration, internal penetration and illegal behavior [2, 3].

Intrusion detection (ID) could find the behaviors of the system which violates the security policy or compromises system security through the system audit data collection, including the operating system, system program, application, and network packet information. On the other hand, Intrusion detection systems could also identify all
kinds of intrusions, including attempting intrusions, ongoing intrusions or the intrusions which has been occurred, and then adopt corresponding protection measures [4]. System with intrusion detection function is called the intrusion detection system. The roles of intrusion detection system are mainly reflected in the following aspects:

(a) Identifying intruders
(b) Identifying intrusion behaviors
(c) Safety breakthrough successful detection and monitoring;
(d) Fighting the intrusion and providing important information in time, and preventing events to expand continuously;
(e) Letting the system come back to the normal state and collect evidences.

The main innovations of this paper lie in the following aspects:

(a) The proposed intrusion detection system is made up five main modules, including Data source, AAA protocol module, FSVM module, Guest computer module and Terminal module.
(b) The intrusion detection module is made up of four modules, which are data gathering module, data preprocessing module, intrusion detecting module and decision response module.
(c) The proposed intrusion detection algorithm is designed utilizing fuzzy support vector machine, and is implemented by training process and testing process.
(d) Fighting the intrusion and providing important information in time, and preventing events to expand continuously;
(e) Letting the system come back to the normal state and collect evidences.

The rest of the paper is organized as follows sections. Section II introduces the related works of this paper and the overview of fuzzy support vector machine. In section III, intrusion detection system based on fuzzy support vector machine is illustrated. In section IV, a series of experiments are conducted to test the effectiveness of the proposed method. Finally, we conclude the whole paper in section V.

II. RELATED WORK

In this section, we will introduce the related works of the proposed paper in two aspects, which are 1) network intrusion detection and 2) the applications of fuzzy support vector machine.

Firstly, we will give some methods for network intrusion detection as follows.

Gil et al. described the design of a collaborative intrusion detection network (CIDN) which is capable of building and sharing collective knowledge about isolated alarms in order to efficiently and accurately detect distributed attacks. It has been also strengthened with a reputation mechanism aimed to improve the detection coverage by dropping false or bogus alarms that arise from malicious or misbehaving nodes. This model will enable a CIDN to detect malicious behaviors according to the trustworthiness of the alarm issuers, calculated from previous interactions with the system [5].

Shakshuki et al. proposed and implemented a novel intrusion-detection system named Enhanced Adaptive ACKnowledge (EAACK) specially designed for MANETs. Compared to contemporary approaches, EAACK demonstrates higher malicious-behavior-detection rates in certain circumstances while does not greatly affect the network performances [6].

Mitchell et al. developed a mathematical model to assess the survivability property of a MCPS subject to energy exhaustion and security failure. The proposed model-based analysis revealed the optimal design setting for invoking intrusion detection to best balance energy conservation versus intrusion tolerance for achieving high survivability. The authors tested the effectiveness of our approach with a dynamic voting-based intrusion detection technique leveraging sensing and ranging capabilities of mobile nodes in the MCPS and demonstrate its validity with simulation validation [7].

Lu et al. proposed an unified detection approach to integrate misuse detection and anomaly detection to overcome their disadvantages. GNP-based class association rule mining method extracts an overwhelming number of rules which contain much redundant and irrelevant information. Particularly, an efficient class association rule-pruning method is proposed based on matching degree and genetic algorithm (GA). In the first stage, a matching degree-based method is applied to preprune the rules in order to improve the efficiency of the GA. In the second stage, the GA is implemented to pick up the effective rules among the rules remaining in the first stage [8].

In paper [9], the authors analyzed the characteristics of current cloud computing, and then proposed a comprehensive real-time network risk evaluation model for cloud computing based on the correspondence between the artificial immune system antibody and pathogen invasion intensity. The paper also combined assets evaluation system and network integration evaluation system, considering from the application layer, the host layer, network layer may be factors that affect the network risks [9].

Xie et al. present a fuzzy clustering algorithm for intrusion detection based on heterogeneous attributes. Firstly, the algorithm modified the comparability measurement for the categorical attributes according to the formula of Hemingway; then, for the shortages of fuzzy C-means clustering algorithm: initialize sensitively and easy to get into the local optimum, the presented new algorithm is optimized by GuoTao approach [10].

Gomez et al. proposed a novel Pareto-based multi-objective evolutionary algorithm which is used to optimize the automatic rule generation of a signature-based intrusion detection system (IDS). This optimizer, included within a network IDS, has been evaluated using a benchmark dataset and real traffic of a Spanish university [11].

Gowrison et al. present an intrusion detection system which is designed to classify by the incorporation of enhanced rules as learnt from the network behavior with less computational complexity of O(n). The proposed method demonstrated the achievements of promising classification rate [12].

Wang et al. analyzed the problem of intrusion detection in a Gaussian-distributed Wireless sensor networks by
characterizing the detection probability with respect to the application requirements and the network parameters under both single-sensing detection and multiple-sensing detection scenarios. Effects of different network parameters on the detection probability were examined in detail [13].

Mitrokotsa et al. examined how to properly use classification methods in intrusion detection for MANETs. In order to do so, the authors evaluated five supervised classification algorithms for intrusion detection on a number of metrics. They measured their performance on a dataset, described in this paper, which includes various traffic conditions and mobility patterns for multiple attacks [14].

Patel et al. surveyed the latest developed IDPSs and alarm management techniques by providing a comprehensive taxonomy and investigating possible solutions to detect and prevent intrusions in cloud computing systems. Considering the desired characteristics of IDPS and cloud computing systems, a list of germane requirements was identified and four concepts of autonomic computing self-management, ontology, risk management, and fuzzy theory are leveraged to satisfy these requirements [15].

Next, as the fuzzy support vector machine is a powerful statistical classification technique based on the idea of structural risk minimization, we will introduce the related works about applications of fuzzy support vector machine as follows.

In paper [16], a fuzzy support vector machines control strategy based on sliding mode control was designed to reduce the oscillation of the sliding mode control. Parameters of fuzzy support vector machines controller were optimized by hybrid learning algorithm, which combines least square algorithm with improved genetic algorithm, to get the optimal control performance for the controlled object. The controller designed consists of a fuzzy sliding mode controller and a fuzzy support vector machines controller, and the compensation controller is selected by comparing switching function with the thickness of boundary layer.

Li et al. proposed a double margin (rough margin) based fuzzy support vector machine (RFSVM) algorithm by introducing rough set into fuzzy support vector machine. Firstly, the authors computed the degree of fuzzy membership of each training sample. Secondly, the data with fuzzy memberships were trained to obtain the decision hyperplane that maximizing rough margin method which contains the lower margin and the upper margin [17].

Chaudhuri et al. [19] utilized a novel Soft Computing tool viz., Fuzzy Support Vector Machine (FSVM) to solve bankruptcy prediction problem. Fuzzy Sets are capable of handling uncertainty and imprecision in corporate data. Thus, using the advantage of Machine Learning and Fuzzy Sets prediction accuracy of whole model is enhanced. FSVM is implemented for analyzing predictors as financial ratios.

From the above related works, we can see that fuzzy support vector machine is a powerful computing tool, and it can be effectively used in network intrusion detection.

As is well known that support vector machine is a useful machine learning method to solve classification problems. However, there are some drawbacks for SVM in many application fields. From the point view of the training set and equations, each training sample should be belonged to either one class or the other, that is, for a given class, it can be easily known that all training samples of a given class are regarded uniformly in the theory of support vector machine. For many real-world application fields, the effects of the training samples may be different for each other. From the above analysis, we can see that each training sample is no more exactly belonged to one of the two classes. For example, a sample in training dataset could possibly 75% belong to one class and 25% to other classes. In the following, we will demonstrate the formal description of fuzzy support vector machine as follows.

In order to modify the importance of training data, we should set a fuzzy membership to each data element. Supposing that we are given a set \( S \) of labeled training data with related fuzzy membership scores as follows.

\[
(y_1, x_1, s_1), (y_2, x_2, s_2), \ldots, (y_l, x_l, s_l)
\]

For each training data sample \( x_i \), a label \( y_i \in \{-1,1\} \) is used as a label for \( x_i \). Particularly, the fuzzy membership \( s_i \) satisfies the condition that \( s_i \in [\delta, 1] \), \( i = 1, 2, \ldots, l \), and the parameter \( \delta \) is larger than zero. As the fuzzy membership \( s_i \), the attitude of data sample \( x_i \), the optimal hyperplane solving problem can be converted as the following equation as follows.

\[
\min \frac{w \cdot w}{2} + C \sum_{i=1}^{l} s_i \cdot \varphi_i
\]

S.T. \( y_i (w \cdot z_i + b) \geq 1 - \varphi_i , i \in \{1,2,\ldots,l\} \), \( \varphi_i \geq 0 \)

In Eq. 2, the parameter \( C \) is a constant. To make the optimization problem more easy to solve, the optimization problem in Eq.2 could be simplified as the following form.

\[
\min W(\lambda) = \sum_{i=1}^{l} \lambda_i - \frac{1}{2} \sum_{i=1}^{l} \sum_{j=1}^{l} \lambda_i \cdot \lambda_j \cdot y_i \cdot y_j \cdot K(x_i, x_j)
\]

S.T. \( \sum_{i=1}^{l} \lambda_i \cdot y_i = 0 \), \( \lambda_i \in [0, s_i \cdot C] \), \( i \in \{1,2,\ldots,l\} \)

Afterwards, Kukn Tucker conditions are used in fuzzy support vector machine, which is illustrated as follows.

\[
\varphi_i (y_i (w \cdot z_i + y) - 1 + \varphi_i) = 0, i \in \{1,2,\ldots,l\}
\]

\[
\varphi_i (s_i C - \lambda_i) = 0, i \in \{1,2,\ldots,l\}
\]
Particularly, the parameter \( C \) is used to modify the maximization of FSVM’s margin and the number of false classification cases. If the number of parameter \( C \) is set larger, the number of false classification cases is smaller and the classifying margin is narrower.

III. INTRUSION DETECTION SYSTEM BASED ON FUZZY SUPPORT VECTOR MACHINE

A. Framework of Intrusion Detection System

As is shown in Fig.1, the framework of the proposed intrusion detection system is illustrated. Intrusion Detection System can detects and analyzes the information which are gathered from a given host machine. On the other hand, intrusion detection system can also detect the intrusion behavior for the machine through collecting information, including file system used, network events, system calls, and so on. Furthermore, Intrusion detection system monitors the modification information in host kernel, host file system and program behavior. The efficiency of Intrusion detection system relies on chosen system characteristics to monitor. Then, each intrusion detection system can find the intrusion behavior in time. From Fig.1, we can see that our intrusion detection system is made up five main modules, which are 1) Data source, 2) AAA protocol, 3) FSVM module located in local computer, 4) Guest computer and 5) Terminals. Then, some detailed analyses are given as follows.

Module 1 is made up of database and storage center, which can store the base data of the proposed system. Particularly, this data source module can store users’ related information and private data.

Module 2 refers to the Authentication, authorization and accounting Protocol, which can detect users’ authentication information. Once a given user is authenticated, this protocol can obtain the specific user’s anomaly level. Afterwards, this protocol can select a specific intrusion detection system which is satisfy user’s anomaly level.

Module 3 is the FSVM module located in local computer, and this module is the key component of our proposed system. We will illustrate this module in the next subsection in detail.

Module 4 and Module 5 and the module utilized by users to access this intrusion detection system.

B. Illustration

Based on the framework of intrusion detection system given above, of which the intrusion detection module based on FSVM is described in this subsection. The proposed intrusion detection module combines anomaly intrusion detection with misuse intrusion detection. Our proposed intrusion detection module is made up of four sections, which are 1) data gathering section, 2) data pre-processing section, 3) intrusion detecting section and 4) decision response section. Particularly, the intrusion detection section includes the misuse intrusion detecting part and the anomaly intrusion detection part. The above two parts are constructed by fuzzy support vector machine, and the organization of our proposed intrusion detection module is shown in Fig.2.

In the following parts, we will give the intrusion detection algorithm based on fuzzy support vector machine.

Firstly, the kernel function \( K() \) of Eq.3 should be determined as follows.

\[
K(x_i, x_j) = \varphi(x_i) \cdot \varphi(x_j)
\]

(6)

where \( \varphi(x) \) a nonlinear function which can map \( x \) to a high dimensional feature space to solve the classification problem. In the proposed intrusion detection model, the kernel function is chosen RBF kernel as follows.
Afterwards, the proposed network intrusion detection model can be implemented by the following equation:

$$f(x) = \text{sign} \left( \sum_{i=1}^{N} \lambda_i y_i K(x_i, x_j) + b \right)$$  \hspace{1cm} (8)$$

In Eq.8 the parameter $b$ is computed by the Karush Kuhn Tucker conditions.

**Algorithm 1:** Intrusion detection algorithm based on fuzzy support vector machine

**Input:** Initial parameters and the kernel function $K()$

**Output:** Network intrusion detection results.

/*Training process*/
(1) Constructing the Training Data of network intrusion detection
(2) Determine the kernel parameter and the trade-off parameter
(3) Select the membership parameter to determine the fuzzy membership of every trial in fuzzy support vector machine
(4) Extract the features of each sample in the training dataset
(5) Train the FSVM classifiers and solve the decision functions

/*Testing process*/
(6) For a given sample
(7) Extract the features of the given sample in the testing dataset
(8) Fuzzy support vector machine classifier by Eq.8
(9) Return the network intrusion detection results

IV. EXPERIMENTS

A. Dataset

In our experiments, the KDD dataset is used which is the common data set used in Intrusion Detection System research papers, and 10 percent of the KDD 99 dataset is used in the propose paper. KDD 99 dataset is utilized for The Third International Knowledge Discovery and Data Mining Tools Competition, which was held in conjunction with KDD-99, The Fifth International Conference on Knowledge Discovery and Data Mining. The competition task was to build a network intrusion detector, a predictive model capable of distinguishing between “bad” connections, called intrusions or attacks, and “good” normal connections. This database contains a standard set of data to be audited, which includes a wide variety of intrusions simulated in a military network environment.

As is shown in paper [20], The KDD Cup 1999 contained 4,898,431 and 311,029 records in the training set and test set, respectively. Each record in our dataset has a class identifier, which can denote either normal or some specific attack class. There are 22 attack classes, which can be classified into four categories, which are DoS, U2R, R2L, and Probe. For the training set, only 19.85% (972,781 records) were normal traffic and the remaining were attack traffic. For the test set, 19.48% (60,593 records) were normal traffic and the remaining were attack traffic. Each record in the KDD Cup 1999 data set contained 41 various quantitative and qualitative features.

B. Experimental Results

To verify the effectiveness of intrusion detecting by our proposed method, some other methods are used to make performance comparison, which are 1) standard SVM, 2) simple incremental SVM (SI-SVM) [21] and 3) peer to peer incremental SVM (P2P-ISVM) [22].

Firstly, we construct a test set which includes more than 90% attacks of new classes (see Table 1).

<table>
<thead>
<tr>
<th>Category name</th>
<th>Number of new attacks</th>
<th>Number of attacks</th>
<th>Ration of new attacks(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DoS</td>
<td>3245</td>
<td>3312</td>
<td>97.98</td>
</tr>
<tr>
<td>U2R</td>
<td>48</td>
<td>48</td>
<td>100</td>
</tr>
<tr>
<td>R2L</td>
<td>521</td>
<td>527</td>
<td>98.86</td>
</tr>
<tr>
<td>Probe</td>
<td>2136</td>
<td>2259</td>
<td>94.56</td>
</tr>
</tbody>
</table>
According to the setting in Table I, the detection rates of each method on the attacks of all categories are obtained, the experimental results are shown in Table II.

<table>
<thead>
<tr>
<th>Method</th>
<th>Normal</th>
<th>Probe</th>
<th>DoS</th>
<th>U2R</th>
<th>R2L</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVM</td>
<td>91.27</td>
<td>64.79</td>
<td>44.12</td>
<td>36.81</td>
<td>17.88</td>
</tr>
<tr>
<td>SI-SVM</td>
<td>98.31</td>
<td>82.34</td>
<td>51.23</td>
<td>48.87</td>
<td>22.74</td>
</tr>
<tr>
<td>P2P-ISVM</td>
<td>99.12</td>
<td>94.12</td>
<td>67.28</td>
<td>54.23</td>
<td>30.81</td>
</tr>
<tr>
<td>Our approach</td>
<td>99.21</td>
<td>98.27</td>
<td>85.23</td>
<td>67.22</td>
<td>38.45</td>
</tr>
</tbody>
</table>

From Table II, we can see that for the detection ration metric, our approach outperforms than other three methods for each category.

Next, the experiments designed to test the performance for each method under the metric of “Correlation Coefficient”, “Detection Rate” and “False Alarm Rate”. The data used in this experiment is collected from KDD 99 dataset, and we separate the whole data into two classes, including training set and test set. For the training dataset, ten sub-dataset are chosen to make performance evaluating, which are named S1 to S10. Each sub-dataset includes 200 normal samples and 200 abnormal samples. Afterwards, we select the normal and attack records with the same number to construct the test set from the test subset. The performance of “Correlation Coefficient (CC)”, “Detection Rate (DR)” and “False Alarm Rate (FAR)” for each method is shown as follows.

The more detailed experimental data of the above are listed in Table III as follows.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Our approach</th>
<th>SI-SVM</th>
<th>P2P-ISVM</th>
<th>SVM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CC</td>
<td>DR</td>
<td>FAR</td>
<td>CC</td>
</tr>
<tr>
<td>S1</td>
<td>0.710</td>
<td>0.753</td>
<td>0.040</td>
<td>0.710</td>
</tr>
<tr>
<td>S2</td>
<td>0.738</td>
<td>0.751</td>
<td>0.041</td>
<td>0.675</td>
</tr>
<tr>
<td>S3</td>
<td>0.764</td>
<td>0.755</td>
<td>0.037</td>
<td>0.657</td>
</tr>
<tr>
<td>S4</td>
<td>0.772</td>
<td>0.796</td>
<td>0.044</td>
<td>0.662</td>
</tr>
<tr>
<td>S5</td>
<td>0.813</td>
<td>0.801</td>
<td>0.042</td>
<td>0.712</td>
</tr>
<tr>
<td>S6</td>
<td>0.821</td>
<td>0.818</td>
<td>0.037</td>
<td>0.696</td>
</tr>
<tr>
<td>S7</td>
<td>0.835</td>
<td>0.832</td>
<td>0.041</td>
<td>0.777</td>
</tr>
<tr>
<td>S8</td>
<td>0.841</td>
<td>0.836</td>
<td>0.037</td>
<td>0.775</td>
</tr>
<tr>
<td>S9</td>
<td>0.848</td>
<td>0.875</td>
<td>0.044</td>
<td>0.805</td>
</tr>
<tr>
<td>S10</td>
<td>0.850</td>
<td>0.882</td>
<td>0.045</td>
<td>0.830</td>
</tr>
</tbody>
</table>
V. CONCLUSIONS

This paper studies automatically detect the network intrusion behavior utilizing fuzzy support vector machine. The proposed intrusion detection system is organized as five modules: 1) Data source, 2) AAA protocol, 3) FSVM module, 4) Guest computer and 5) Terminals. The intrusion detection module is made up of four sections, which are 1) data gathering section, 2) data pre-processing section, 3) intrusion detecting section and 4) decision response section. Afterwards, the intrusion detection algorithm based on fuzzy support vector machine is designed by training process and testing process.

ACKNOWLEDGMENT

This work is supported by the Science and Technology of Hunan Province (2012GK3195) and the Youth Foundation of educational science of Hunan Province “12th Five-Year Plan” (XJK011QZJ011).

REFERENCES


